What is 'Cyber Security'?: Differential Language of Cyber Security Across the Lifespan

Simon L. Jones
Dept. of Computer Science
University of Bath
Bath, UK
S.L.Jones@bath.ac.uk

Emily I. M. Collins
School of Management
University of Bath
Bath, UK
E.I.M.Collins@bath.ac.uk

Ana Levordashka
School of Management
University of Bath
Bath, UK
A.Levordashka@bath.ac.uk

Kate Muir
School of Management
University of Bath
Bath, UK
K.Muir@bath.ac.uk

Adam Joinson
School of Management
University of Bath
Bath, UK
A.Joinson@bath.ac.uk

ABSTRACT

People experience and understand cyber security differently. Our ongoing work aims to address the fundamental challenge of how we can understand a diverse range of cyber security experiences, attitudes and behaviours in order to design better, more effective cyber security services and educational materials. In this paper, we take a lifespan approach to study the language of cyber security across three main life stages - young people, working age, and older people. By applying text feature extraction and analysis techniques to lists of cyber security features generated by each age group, we

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5971-9/19/05…$15.00
https://doi.org/10.1145/3290607.3312786

Table 1: Survey Respondents.

Sample	N	Mean Age (S.D.)	Gender (% Female)
Children	146	13.16 (1.03)	56.17
Working Age	211	36.34 (4.25)	74.41
Older Adults	146	71.73 (6.27)	34.93

The child sample were recruited primarily from two secondary schools, located in the South West and North East of England. Working age adults were recruited through social media, e-mail lists, community groups and a participant panel, Prolific. Older adults were recruited through community groups, existing research participant databases, and organisations working with older adults.

Participants were presented with instructions to list characteristics or features they think of when they hear the term 'cyber security', and to consider how they would explain it to someone who did not know what it meant. They were assured that there were no right or wrong answers, and told to include obvious answers, as well as both positive and negative aspects if possible. To further clarify the task, participants were also given an example of the kinds of features one might suggest if asked to do the same for the concept of 'love' (e.g. caring, trust, affection, friendship).

Sidebar 1: Recruitment and Survey Instructions

illustrate the differential language of cyber security across the lifespan and discuss the implications for design and research in HCI.

KEYWORDS

Cyber Security; Security; Lifespan; Ageing; Text Analysis; LIWC; TF-IDF;

INTRODUCTION

The pervasive and constantly evolving nature of technology means that cyber security incidents stand to impact more people than ever before. Users are often described as the weakest link in security systems [15], hence it is important to understand factors that may influence their security-related behaviours and decisions. This aim is complicated by *differential vulnerabilities* in cyber security [10]; that is, the notion that cyber security can differ between individuals, and be *"socially contingent and differentially applied"* based on factors such as gender, age and experience. Age in particular is an important factor when considering individual responses to technological and social change, not least because different life stages bring distinct and diverse social, organisational and environmental contexts and challenges [11].

Certain features of cyber security may be more salient to individuals at different life stages. For example, children in some schools receive education in cyber security as part of the curriculum. Their experience with instances of cyber security could therefore be different from the experiences of working age and older adults, who may be less likely to be formally educated about the subject. Similarly, older adults who are retired might encounter different varieties of cyber security incidents compared to working age adults in the workplace. Therefore, attempts to understand what cyber security *means* to technology users need to consider these potential differences. This is not only to inform the design of more effective security interfaces and educational materials, but also to better comprehend how different contexts and experiences impact their understanding.

The work presented in this paper investigates differences in the *language* of cyber security across three main life stages - young people, working age, and older adults. We present a survey designed to capture the language that people in each life stage use to describe features of cyber security. We apply text analysis methods to the survey results in order to identify similarities and differences in the language of cyber security between age groups, and a discuss the possible implications for design and research in human-computer interaction and cyber security.

RELATED WORK

Previous work has highlighted substantial age-related differences in cyber security behaviours and risks. For example, young people are most likely to be subject to cyber-bullying and grooming [6], more susceptible to phishing than other age groups [14], and more likely to share passwords [16]

a) Children

b) Working Adults

c) Older Adults

Figure 1: Word clouds for the most frequently listed features of cyber security for each of the three life stages. Larger items indicate more frequent features.

and other personal information on social media sites [2]. Older adults are less likely to adopt PIN or biometric protections for their devices [4], and have been found to be more susceptible to reciprocity-based *weapons of influence*, whereby adversaries offer rewards for compromising security [8] (e.g. installing malware when lured with a free gift). Older adults have also been found to engage in fewer privacy-related behaviours on social networking sites, but also disclose less information online, and express more concern about other people's privacy, than younger age groups [5].

The perceived personal relevance of cyber security information and threats has an impact on individuals propensity to report security incidents [1]. Hence, the varying ways in which groups understand, define and discuss cyber security must become a key design concern for interactive systems, security warnings, training tools and educational materials. Relatively little work has taken a lifespan approach to studying cyber security, or explored how differences in cyber security related behaviours are reflected in, or even caused by, differences in the meaning of the term 'cyber security' to different age groups.

METHOD: CYBER SECURITY FEATURES SURVEY

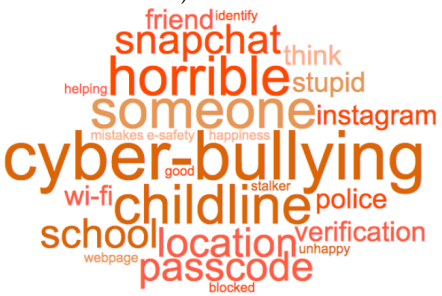
We conducted a survey in order to generate a list of features that people associated with the term 'cyber security'. Respondents from a sample of children, adults of working age and older adults were asked to list characteristics of cyber security (see Sidebar 1 for recruitment details and instructions given to survey respondents). A total of 503 respondents completed the survey. This comprised children aged from 11-18 (n=146), working age adults (n=211), and older adults aged 60 or over (n=146). Further demographics for each sample can be found in Table 1. Across the three samples, a total of 2897 cyber security features were provided. Each participant reported an average of 5.79 features (SD=4.30). Older adults reported more features on average (M=6.89, SD=4.64), followed by working age participants (M=5.30, SD=4.16). Children reported the fewest features (M=4.07, SD=3.53).

Two independent judges extracted individual features from the responses (i.e. if a participant wrote a full sentence rather than a word or a short phrase, it was divided into individual features where appropriate). All stop-words (common words and short function words used in everyday language e.g. *the*, *is*, *at*, *which*, *and*, *of*) were removed from the dataset. Spelling checks were applied to correct misspelled words. The features were also processed in order to group minor variations of the same root word (e.g. hack, hacks, hacked). We then performed three types of analysis on the lists of features generated by each life stage group: (1) a basic frequency analysis (identifying the most commonly occurring features), (2) TF-IDF feature extraction and (3) LIWC dictionary analysis.

TF-IDF Feature Extraction

TF-IDF is a technique that assesses the frequencies of terms in a dataset as they compare to frequencies within subsets of the data [12], which in this case refers to the three life stages. The technique scores

a) Children



b) Working Adults

c) Older Adults

Figure 2: Word clouds for TF-IDF extracted features for each life stage. Larger items in the word cloud indicate more informative (higher TF-IDF scoring) features.

terms according to both their frequency within a life stage (Term Frequency), and their uniqueness to that life stage (Inverse Document Frequency). A TF-IDF score was calculated for each feature.

LIWC Dictionary Analysis

We also analysed the features listed by each life stage group using the LIWC2015 text analysis tool [9]. LIWC compares each word in a text against a set of validated dictionaries associated with psychologically-relevant categories (e.g. cognitive processes, core drives and needs, positive and negative emotions, personal concerns) and calculates the percentage of total words that match each of the categories.

A MANOVA analysis with post-hoc Tukey HSD comparisons (see Sidebar 2) was performed in order to identify LIWC categories for which occurrences differed between life stages.

RESULTS

Feature Frequency and TF-IDF Analysis Results

Figure 1 shows three word clouds illustrating the most frequently listed features of cyber security for each age group. Features such as *passwords*, *safety*, *protection*, *internet*, *virus*, *computer* and *hacking* were highly prevalent across all three age groups. This suggests that there are central features associated with a universal prototype of cyber security (i.e. aspects of cyber security which are salient across the lifespan).

However, the TF-IDF analysis reveals a set of salient cyber security features that distinguish the three life stages (see Figure 2). Children were unique in listing terms relating to *cyber-bullying*, *e-safety*, *strangers*, *friends* and social media services such as *Snapchat* and *Instagram*. Working age adults generated features that were often technically-focused (e.g. *authentication*, *encryption*, *https*, *penetration*, *VPN*) or which focused on criminal aspects of cyber security (e.g. *criminals*, *terrorism*, *stealing*). Older adults were relatively unique in associating cyber-security with features such as *control*, *intrusion*, *nuisance*, *malware*. *trolling*, *society* and *possessions*.

LIWC Analysis Results

The LIWC analysis (see Sidebar 2 for statistical analysis and Figure 3 for graphs) revealed significant differences ($p<0.05$) in: *positive emotion* features between children (M=15.2, S.D.=19.4) and older adults (M=11.8, S.D.=19.9); *anxiety* related features between children (M=0.2, S.D.=1.5) and both working age (M=2.3, S.D.=10.7) and older adults (M=2.2, S.D.=6.5); *social processes* between children (M=9.0, S.D.=13.6) and working age adults (M=4.7, S.D.=13.5) ; *cognitive processes* between older adults (M=7.9, S.D.=9.9) and both children (M=3.6, S.D.=7.2) and working age adults (M=4.2, S.D.=11.3); *space* related features between older adults (M=3.1, S.D.=5.2) and both children (M=1.6, S.D.=4.4) and working age

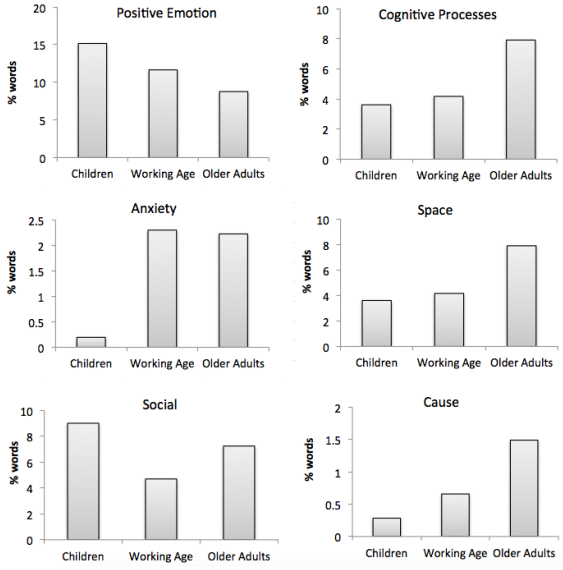


Figure 3: Graphs comparing the percentage of words from LIWC dictionary categories.

A MANOVA was conducted to assess statistically significant differences in LIWC category scores across the three life stages, $F(160, 1018)=1.633$ $p < .05$; Wilk's $\Lambda=0.633$, partial-$\eta^2=.204$.

Post-hoc Tukey HSD tests (adjusted for multiple comparisons) revealed which LIWC categories are affected by life stage. Graphs above show categories for which statistically significant differences were detected: *positive emotion, anxiety, social, cognitive processes, space, cause.*

Sidebar 2: Analysis of LIWC word frequencies.

adults (M=1.5, S.D.=4.8); and *cause* related features between children (M=0.3, S.D.=1.6) and older adults (M=1.5, S.D.=3.6).

DISCUSSION AND CONCLUSION

This study explores the ways in which children, working age adults and older adults describe cyber security. The TF-IDF analysis identifies differences in the salience of features for the three groups. For example, children focus more on social interactions and processes, and predatory/bullying behaviour, and adults focus more on malicious acts and technical protection measures. Our LIWC analysis highlights differences in the language used by each group, for example suggesting that children exhibit more positive emotion and less anxiety than adults when describing cyber security. These differences may reflect the divergent advice given to, and events experienced by, these groups, shaping their concept of cyber security in different ways. Future work may wish to explore how these conceptualisations are formed, and more closely examine how they impact behaviours.

Existing work in this area has focused on studying differences in cyber security-related *behaviours*, rather than conceptualisation and understanding. For example, [7] revealed a 'U-shaped curve' whereby the youngest and oldest members of society are less protective than the middle-aged cohort. Our findings offer further support for age-related differences, suggesting that behavioural differences occur alongside varying conceptualisations of what cyber security entails.

Furthermore, uncovering variations in the language of cyber security also has implications for education and communications. For instance, relating advice to an age group's salient concerns (e.g. social media and cyber-bullying for children, banking for working age, etc.), or explicitly widening explanations of cyber security to include unfamiliar features, might support education. Moreover, whilst the differences observed appear to align with specific risks each age group are likely to face, an important consideration might be how well these age groups are prepared for life stage transitions (e.g. from childhood to adulthood, or working age to retirement), and whether they have an awareness of other aspects of cyber security that may become more relevant?

Our work also highlights important differences that may be relevant to those designing computerised systems to detect and signal potential cyber security issues (e.g. mining online chatter to support cyber threat warning systems, as in [13]). Previous work at CHI has presented dictionary-based text analysis tools for studying key HCI concerns (such as privacy [3]) in a naturalistic way. Our ongoing work aims to progress the development of similar dictionary-based tools to support unbiased analysis of cyber security in HCI research.

REFERENCES

[1] Pam Briggs, Debora Jeske, and Lynne Coventry. 2017. The Design of Messages to Improve Cybersecurity Incident Reporting. In *International Conference on Human Aspects of Information Security, Privacy, and Trust*. Springer, 3–13.

ACKNOWLEDGMENTS

The 'Cyber Security Across the LifeSpan' (cSALSA) research project is a collaboration between the University of Bath, Cranfield University, University of Northumbria and University of Portsmouth. It is funded by the EPSRC (EP/P011454/1). We wish to thank all of the participants in our research for their time and cooperation.

[2] Elizabeth B Dowell, Ann W Burgess, and Deborah J Cavanaugh. 2009. Clustering of Internet risk behaviors in a middle school student population. *Journal of School Health* 79, 11 (2009), 547–553.

[3] Alastair J Gill, Asimina Vasalou, Chrysanthi Papoutsi, and Adam N Joinson. 2011. Privacy dictionary: a linguistic taxonomy of privacy for content analysis. In *Proc. of the SIGCHI conference on human factors in computing systems*. ACM, 3227–3236.

[4] Marian Harbach, Alexander De Luca, Nathan Malkin, and Serge Egelman. 2016. Keep on Lockin' in the Free World: A Multi-National Comparison of Smartphone Locking. In *Proc. of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 4823–4827. https://doi.org/10.1145/2858036.2858273

[5] Murat Kezer, Barış Sevi, Zeynep Cemalcilar, and Lemi Baruh. 2016. Age differences in privacy attitudes, literacy and privacy management on Facebook. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 10, 1 (2016).

[6] Sylvia Kierkegaard. 2008. Cybering, online grooming and ageplay. *Computer Law & Security Review* 24, 1 (2008), 41–55.

[7] Linda Little, Pam Briggs, and Lynne Coventry. 2011. Who knows about me?: an analysis of age-related disclosure preferences. In *Proc. of the 25th BCS Conference on Human-Computer Interaction*. British Computer Society, 84–87.

[8] Daniela Oliveira, Harold Rocha, Huizi Yang, Donovan Ellis, Sandeep Dommaraju, Melis Muradoglu, Devon Weir, Adam Soliman, Tian Lin, and Natalie Ebner. 2017. Dissecting spear phishing emails for older vs young adults: On the interplay of weapons of influence and life domains in predicting susceptibility to phishing. In *Proc. of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 6412–6424.

[9] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. *The development and psychometric properties of LIWC2015.* Technical Report.

[10] James Pierce, Sarah Fox, Nick Merrill, and Richmond Wong. 2018. Differential Vulnerabilities and a Diversity of Tactics: What Toolkits Teach Us About Cybersecurity. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 139 (Nov. 2018), 24 pages. https://doi.org/10.1145/3274408

[11] Martin Pinquart and Rainer K Silbereisen. 2004. Human development in times of social change: Theoretical considerations and research needs. *International Journal of Behavioral Development* 28, 4 (2004), 289–298.

[12] Juan Ramos et al. 2003. Using tf-idf to determine word relevance in document queries. In *Proc. of the first instructional conference on machine learning*, Vol. 242. 133–142.

[13] Anna Sapienza, Sindhu Kiranmai Ernala, Alessandro Bessi, Kristina Lerman, and Emilio Ferrara. 2018. DISCOVER: Mining Online Chatter for Emerging Cyber Threats. In *Companion Proc. of the The Web Conference 2018 (WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 983–990. https://doi.org/10.1145/3184558.3191528

[14] Steve Sheng, Mandy Holbrook, Ponnurangam Kumaraguru, Lorrie Faith Cranor, and Julie Downs. 2010. Who Falls for Phish?: A Demographic Analysis of Phishing Susceptibility and Effectiveness of Interventions. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 373–382. https://doi.org/10.1145/1753326.1753383

[15] Ryan West, Christopher Mayhorn, Jefferson Hardee, and Jeremy Mendel. 2009. The weakest link: A psychological perspective on why users make poor security decisions. In *Social and Human elements of information security: Emerging Trends and countermeasures*. IGI Global, 43–60.

[16] Monica Whitty, James Doodson, Sadie Creese, and Duncan Hodges. 2015. Individual differences in cyber security behaviors: an examination of who is sharing passwords. *Cyberpsychology, Behavior, and Social Networking* 18, 1 (2015), 3–7.